Standards for Reuse and Interoperability at the GES DISC

Christopher Lynnes

Goddard Earth Sciences

Distributed Information Services Center

Why Reusable Standards?

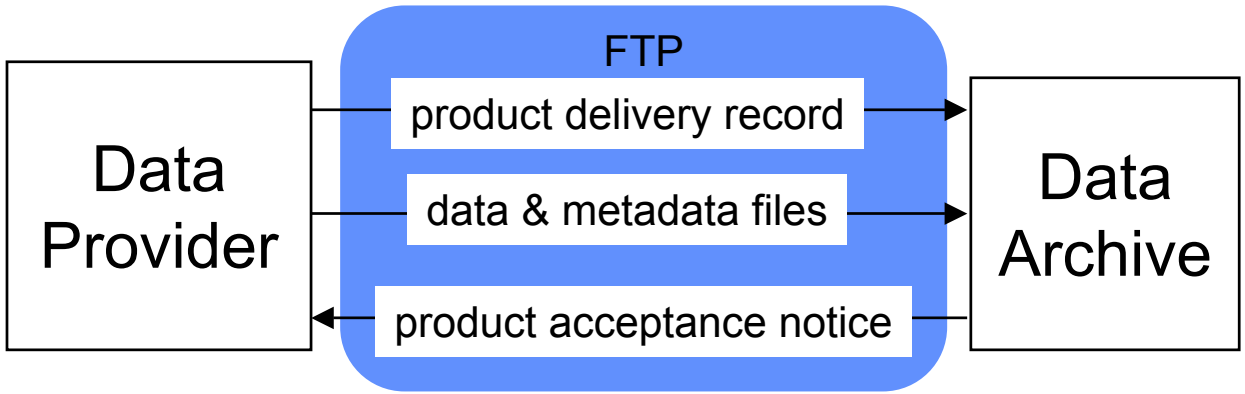
- Interoperability
  - Supports access by external organizations and users
  - But: difficult to gauge utility and usability
- Reuse
  - Serve as specifications
  - Saves on requirements, design, implementation
  - Particularly useful internally for large, complex systems

Reusable Standards

- Data Exchange: PDR/PAN
- Data Services: OPeNDAP
- Data Services: OpenGIS
- Data Model: ECHO
- Data Model: GCMD
- Data Services: Web Services

Data Exchange Standard: PDR/PAN

- Product Delivery Record / Product Acceptance Notice
  - Also known as SIPS interface



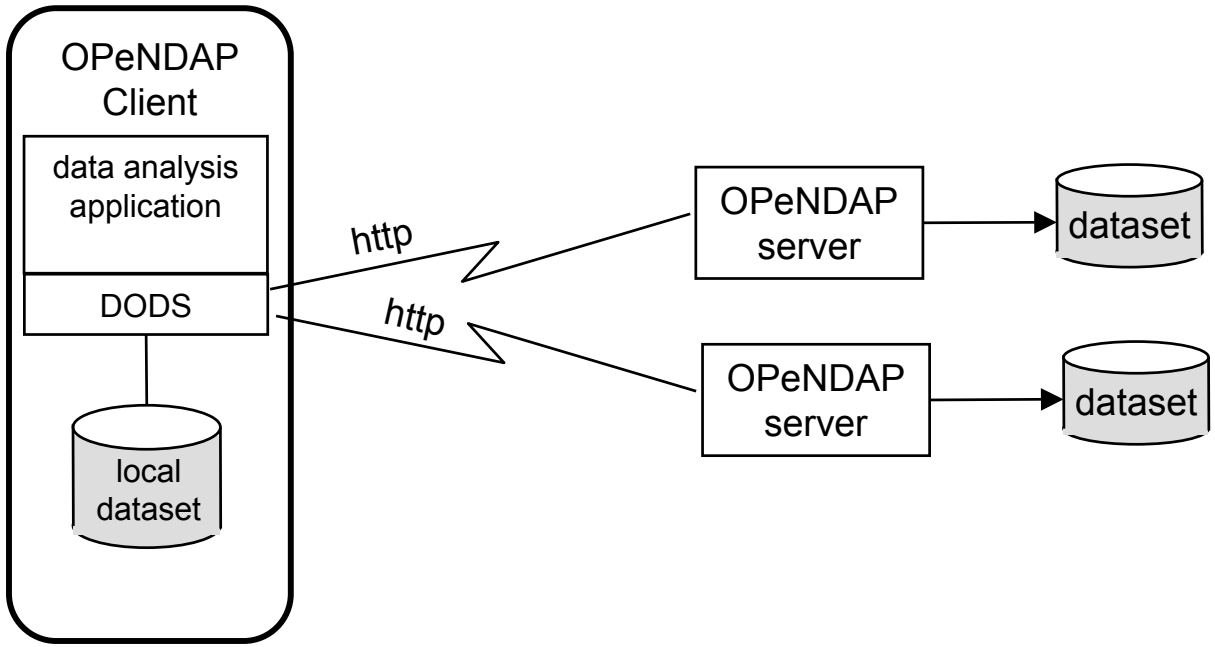
- Key Characteristics
  - Operational
  - Peer-to-peer
  - Asynchronous

PDR/PAN Interface cont.

- Used to exchange MODIS, AIRS, MLS, OMI data
- Reused to shuttle data locally within GES DISC
- Advantages
  - Well-documented (ECS-SIPS ICD, Vol. 0)
  - High performance
- Disadvantage
  - Whole files only
  - PANs do not specify files successfully ingested

Data Services: OpeNDAP

- Open-source Project for a Network Data Access Protocol (OPeNDAP)
  - Formerly known as DODS
  - TRMM, MODIS, AIRS data available

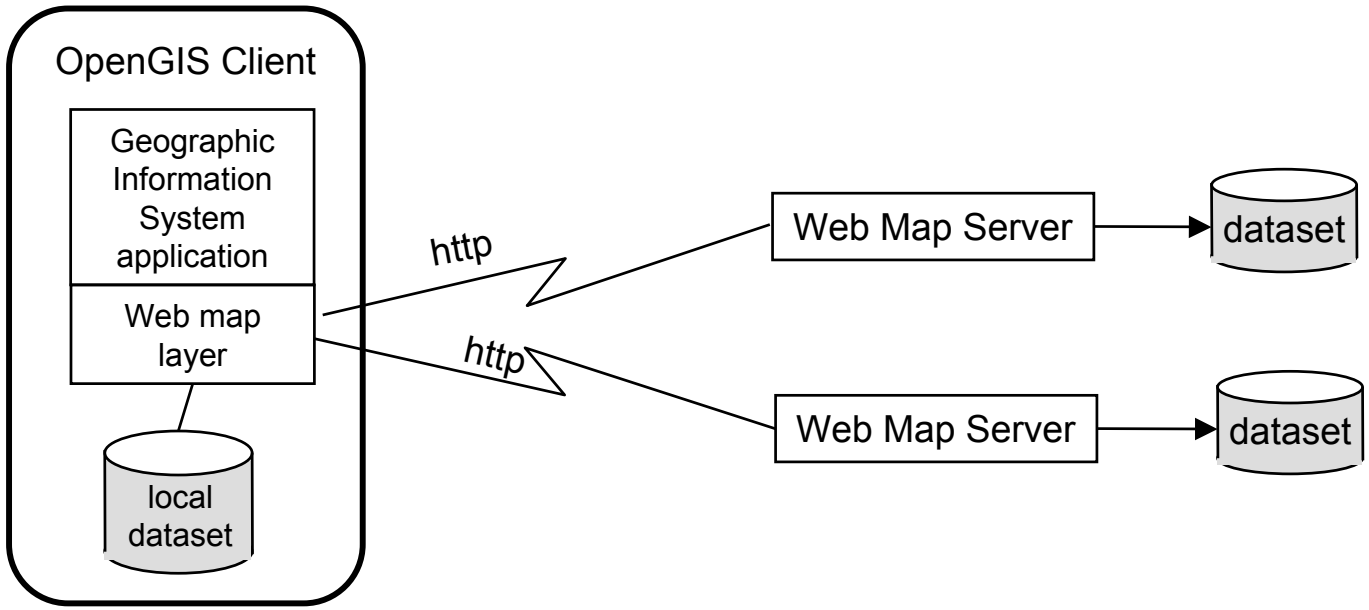


OPeNDAP Trade

- Key Advantages
  - HDF support
  - Server-side subsetting
- Disadvantages
  - Installation can be tricky
  - Cataloguing still evolving
    - GCMD, ECHO, THREDDS, other?
  - Performance overhead of format translations?
- Internal Reuse: intra-system data intercomparison

Data Services: OpenGIS

- Client-server geographic data and pictures Running two servers
  - Univ. Minn. MapServer
  - Synergy Standalone OGC Server
  - Limited data, but more coming



OpenGIS Trade

- Advantages
  - Regular georeferenced data are ***much*** easier for users and clients to use
- Disadvantages
  - Takes a fair bit of tailoring
  - Not friendly to time, depth/height dimensions
  - Cataloguing still evolving
  - Performance of on-the-fly reprojection and rendering
- Dilemmas:
  - Which version of spec?
  - Which server implementation?
  - How do clients see data?

Data Model: ECHO

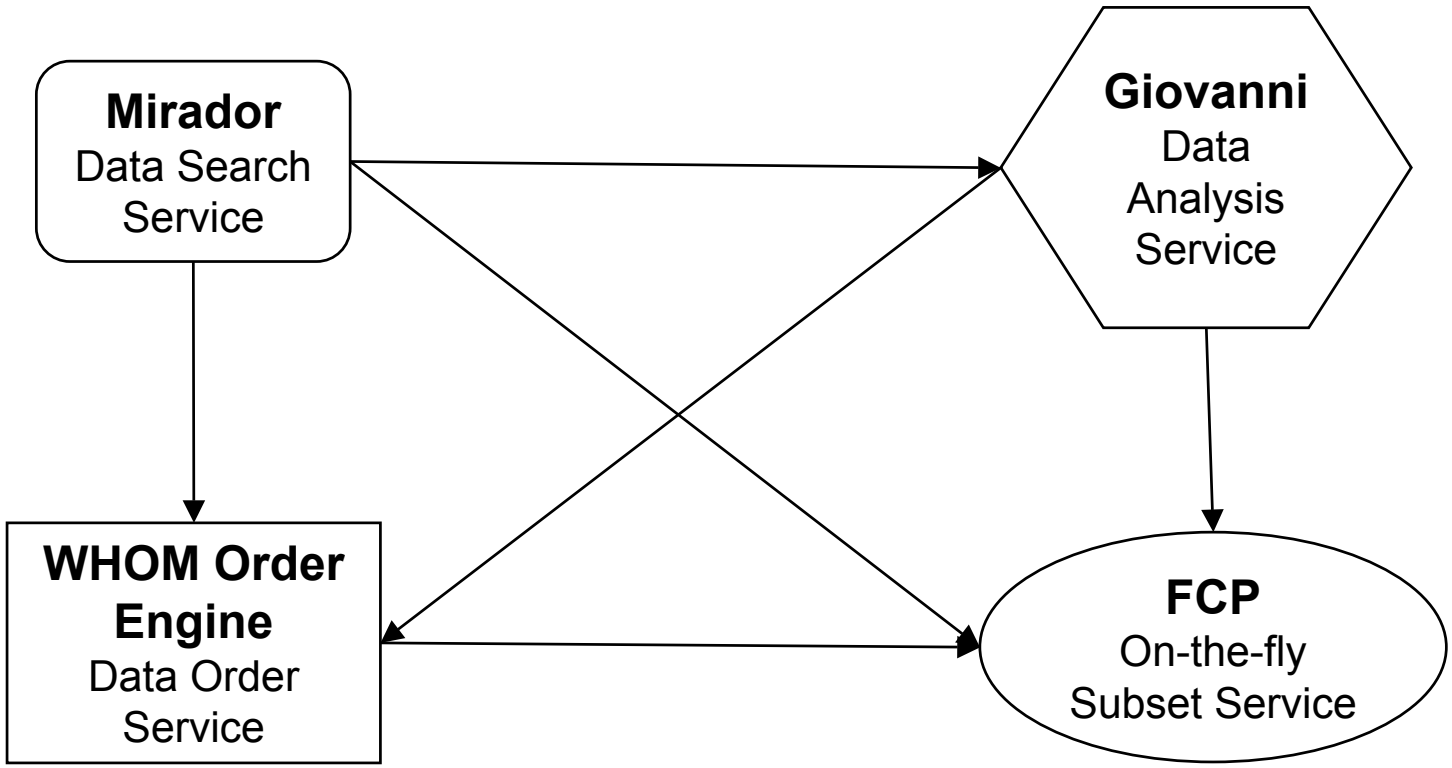
- EOS Clearinghouse (ECHO)
  - Includes collection- and file-level metadata
  - XML-based
  - Data Model based loosely on EOSDIS model
- Reused (sort of) for S4PA
  - Simple, Scalable, Script based Science Processing Archive
    - Disk-based data management and archive system
  - Avoids painful process of defining data model
  - Easy publication of metadata to ECHO
    - Uses XSLT to handle minor deviations from ECHO

GCMD DIFs: Collection-level Metadata

- Data Interchange Format (DIF)
  - Convertible to XML, FGDC
  - Full-featured system (GCMD) for inserting and updating DIFs
- Reuse: GCMD for managing local collection metadata(?)
- Advantages:
  - Easy step to interoperability
- Disadvantages:
  - Cedes some control

Architecture: Web Services

- Based on SOAP, WSDL, UDDI



Web Services Reuse and Interoperability

- Goal: Service-Oriented Architecture
  - Plug and play services
- Internal Reuse
  - Supply services among applications
  - Standardize passing of context: Datasets, parameters, time
- External Interoperability
  - Services available to external organizations/clients
- Dilemmas
  - Language: Perl, java…
  - Which services can we afford to make available externally?

On the Radar: Grid

- Popular way to link distributed systems
  - Could help tie together several mini-systems within GES DISC
- But:
  - Tricky to implement
  - Better on some platforms than other
  - Which version?

Standards Adoption at the GES DISC

- Mandate?
  - E.g. PDR/PAN, GCMD, ECHO
- Benefit to GES DISC User Community?
  - Sometimes cloudy…
- Stability?
  - Backward compatibility is key
- Implementation Cost?
  - Reuse open-source/commercial software?
- Internal Reuse?
  - Amortizes implementation cost